

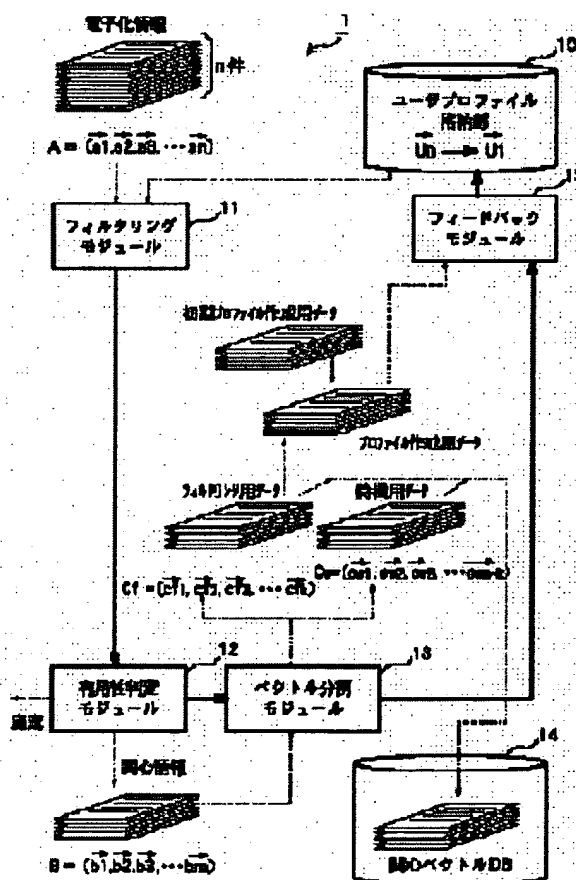
INFORMATION FILTERING METHOD, INFORMATION FILTERING DEVICE, AND RECORDING MEDIUM

Patent number: JP10340270
Publication date: 1998-12-22
Inventor: HARA MASAMI; KITANI TSUYOSHI
Applicant: N T T DATA KK
Classification:
 - international: G06F17/30
 - european:
Application number: JP19970149790 19970606
Priority number(s): JP19970149790 19970606

Report a data error here

Abstract of JP10340270

PROBLEM TO BE SOLVED: To provide an information filtering device that can execute a filtering, in which the concern of a user is precisely reflected. **SOLUTION:** The set of information vectors, whose similarity with a user profile, which shows the concern degree of the user against electronized information that is previously decided exceeds a first reference value, is specified by a filtering module 11 and a usefulness judgment module 12 selects a vector set B with which the user is concerned. The selected vector set B is divided by a vector division module 13 and only the set of the information vectors, whose mutual similarity exceeds a second reference value, is set to be a feedback vector set. A feedback module 15 updates the user profile base don the feedback vector set.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-340270

(43) 公開日 平成10年(1998)12月22日

(51) Int.Cl.⁶

識別記号

F I

G 0 6 F 17/30

G 0 6 F 15/403

3 4 0 A

15/40

3 1 0 F

審査請求 未請求 請求項の数 8 O L (全 9 頁)

(21) 出願番号 特願平9-149790

(22) 出願日 平成9年(1997)6月6日

(71) 出願人 000102728

株式会社エヌ・ティ・ティ・データ
東京都江東区豊洲三丁目3番3号

(72) 発明者 原 正巳

東京都江東区豊洲三丁目3番3号 エヌ・
ティ・ティ・データ通信株式会社内

(72) 発明者 木谷 強

東京都江東区豊洲三丁目3番3号 エヌ・
ティ・ティ・データ通信株式会社内

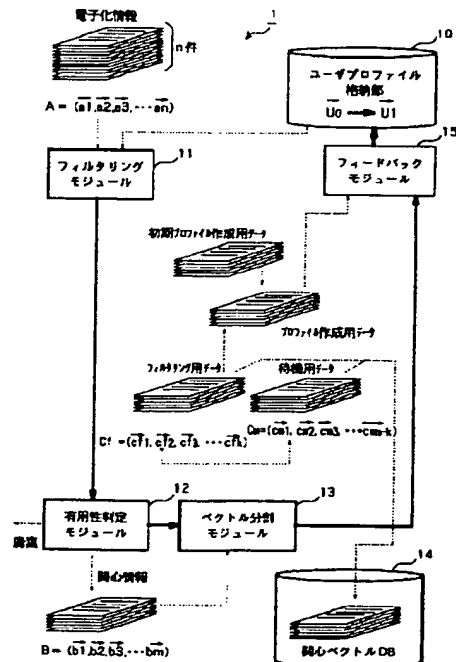
(74) 代理人 弁理士 鈴木 正剛

(54) 【発明の名称】 情報フィルタリング方法、情報フィルタリング装置、及び記録媒体

(57) 【要約】

【課題】 ユーザの関心が正確に反映されるようなフィルタリングを可能にする情報フィルタリング装置を提供する。

【解決手段】 下記の機能を備えて情報フィルタリング装置を構成する。即ち、予め定めた電子化情報に対するユーザの関心度合いを表すユーザプロフィールとの類似度が第1基準値を越える情報ベクトルの集合をフィルタリングモジュール11で特定し、さらに、有用性判定モジュール12で、ユーザが関心をもつベクトル集合Bを選別する。選別されたベクトル集合Bはベクトル分割モジュール13で分割され、互いの類似度が第1基準値よりも大きい第2基準値を越える情報ベクトルの集合のみをフィードバック用ベクトル集合とする。フィードバックモジュール15は、このフィードバック用ベクトル集合に基づいてユーザプロフィールを更新する。



【特許請求の範囲】

【請求項1】 複数の電子化情報をそれぞれ情報ベクトルとして表現し、これらの情報ベクトルのうち、予め定めた電子化情報に対するユーザの関心度合いを表す基準ベクトルとの類似度が第1基準値を越え、且つユーザが有用と判定した情報ベクトルの集合を第1ベクトル集合として特定する過程と、

特定した第1ベクトル集合から互いの類似度が前記第1基準値より大きい第2基準値を越える情報ベクトルの集合を抜き出して第2ベクトル集合とし、この第2ベクトル集合に基づいて前記基準ベクトルを更新する過程とを含む、

更新された基準ベクトルを次の第1ベクトル集合の特定に用いることを特徴とする、コンピュータ装置を用いた情報フィルタリング方法。

【請求項2】 複数の電子化情報をそれぞれ情報ベクトルとして表現し、これらの情報ベクトルのうち予め定めた電子化情報に対するユーザの関心度合いを表す基準ベクトルとの類似度が第1基準値を越え、且つユーザが有用と判定した情報ベクトルの集合を第1ベクトル集合として特定する過程と、

特定した前記第1ベクトル集合と過去に特定された前記第1ベクトルとの和集合から互いの類似度が前記第1基準値より大きい第2基準値を越える情報ベクトルの集合を抜き出して第2ベクトル集合とし、この第2ベクトル集合に基づいて前記基準ベクトルを更新する過程とを含む、

更新された基準ベクトルを次の第1ベクトル集合の特定に用いることを特徴とする、コンピュータ装置を用いた情報フィルタリング方法。

【請求項3】 前記過去に特定された第1ベクトル集合のうち、生成された後所定期間を経過した情報ベクトルを除去したうえで前記第2ベクトル集合を抜き出すことを特徴とする請求項2記載の情報フィルタリング方法。

【請求項4】 それぞれ情報ベクトルとして表現された複数の電子化情報を入力するとともに、これらの情報ベクトルのうち予め定めた電子化情報に対するユーザの関心度合いを表す基準ベクトルとの類似度が第1基準値を越える情報ベクトルの集合を特定するフィルタリング手段と、

このフィルタリング手段により特定された情報ベクトルの集合からユーザが有用と判定した第1ベクトル集合を選別する有用性判定手段と、

前記第1ベクトル集合から互いの類似度が前記第1基準値よりも大きい第2基準値を越える情報ベクトルの集合を第2ベクトル集合として特定するベクトル分割手段と、

前記第2ベクトル集合に基づいて前記基準ベクトルを更新するベクトル更新手段と、を備えて成る情報フィルタリング装置。

【請求項5】 前記ベクトル分割手段は、第1ベクトル集合に属する個々の情報ベクトルについて互いの類似度が前記第2基準値を越える情報ベクトルの数を検出し、検出数が最大になるベクトル集合を前記第2ベクトル集合として特定するように構成されることを特徴とする請求項4記載の情報フィルタリング装置。

【請求項6】 前記ベクトル分割手段は、前記検出数が同数となるベクトル集合が複数存在する場合に時間的に新しい情報ベクトルをより多く含む一つのベクトル集合を第2ベクトル集合として決定するように構成されることを特徴とする請求項5記載の情報フィルタリング装置。

【請求項7】 前記第1ベクトル集合を一定時間蓄積するベクトル蓄積手段をさらに備え、前記ベクトル分割手段は、最新の第1ベクトル集合と前記ベクトル蓄積手段に蓄積されている過去の第1ベクトル集合との和集合から前記第2ベクトル集合を特定するように構成されることを特徴とする請求項6記載の情報フィルタリング装置。

【請求項8】 複数の電子化情報をそれぞれ情報ベクトルに変換する処理、

変換された複数の情報ベクトルのうち、予め定めた電子化情報に対するユーザの関心度合いを表す基準ベクトルとの類似度が第1基準値を越え、且つユーザが有用と判定した情報ベクトルの集合を第1ベクトル集合として特定する処理、

及び、特定した第1ベクトル集合から互いの類似度が前記第1基準値より大きい第2基準値を越える情報ベクトルの集合を抜き出して第2ベクトル集合とし、この第2ベクトル集合に基づいて次の第1ベクトル集合の特定に用いる前記基準ベクトルを更新する処理、

をコンピュータ装置に実行させるための情報フィルタリング用プログラムを当該コンピュータ装置が読み取り可能な形態で記録して成る記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、自然言語処理技術及び情報検索技術に関し、例えば、断続的に流入する大量の電子化情報から、ユーザにとって関心の高いテーマを持つ電子化情報をフィルタリング（選別処理、以下同じ）する手法に関する。

【0002】

【従来の技術】インターネットの普及等により電子化情報の流通の機会が増すにつれて、大量の電子化情報から必要な情報のみを取捨選択する必要が生じてくる。この場合の取捨選択作業は、人手で行うには負担がかかりすぎるため、コンピュータ装置による自動化、例えば、ユーザが関心を持つテーマに沿って、流入する大量の電子化情報を自動的に選別する情報フィルタリングに関する検討がなされている。この情報フィルタリングでは、ユ

ユーザの関心度合いを定量化してコンピュータ処理するために、ユーザがどのような情報に関心を有しているかを表現する基準ベクトル（ユーザプロファイルベクトル、ユーザプロファイル、あるいは単にプロファイルとも呼ばれている）が用いられる。ユーザプロファイルは、例えば、予めユーザが関心のある電子化情報に含まれる複数のテキストデータの集合に含まれる単語の出現頻度（ユーザにとっては重要度を表す）を単語毎に求め、求めた単語の種類に応じた次元（例えば、10種類の単語であれば10次元）のベクトルに変換するとともに、これを正規化したものである。

【0003】また、上記ユーザプロファイルはユーザの関心をできるだけ正確に反映し、フィルタリング精度を高めるために、過去のフィルタリング結果を次回以降のフィルタリングに反映させる適合フィードバックを利用した手法も検討されている。

【0004】図7は、適合フィードバックを利用した従来の情報フィルタリング装置の機能ブロックと、フィルタリング処理の概要を示す図である。図中、実線は処理の流れ、破線はデータ類の流れを表している。この情報フィルタリング装置2は、複数（ n 件）の文書を例えばイメージ処理して電子化情報に変換するとともに個々の電子化情報をそれぞれ該当する情報ベクトルに変換し、 n 件分の情報ベクトルの集合（ベクトル集合） $A = \{a_1, a_2, a_3, \dots, a_n : a_1, a_2, \dots$ は情報ベクトル）を生成する手段（図示省略）を有し、さらに、コンピュータ装置が所定のプログラムを読み込んで実行することにより形成される、ユーザプロファイル格納部10、フィルタリングモジュール11、有用性判定モジュール12、フィードバックモジュール13の各機能ブロックを備えている。

【0005】フィルタリングモジュール11は、ユーザプロファイル格納部10に格納されたユーザプロファイル u （小文字はベクトル表記とする、以下同じ）と上記ベクトル集合 A との類似度が基準値を越える情報ベクトルを選別するものである。有用性判定モジュール12は、フィルタリングモジュール11で選別された情報ベクトルのうち、ユーザが真に自分の関心と合致している、つまり有用性があると判定したものの集合（ベクトル集合） $B = \{b_1, b_2, \dots, b_m : b_1, b_2, \dots$ は情報ベクトル、但し $m \leq n$ ）を選別するものであり、有用性のない情報ベクトルについてはここで破棄する。ベクトル集合 B に属する情報ベクトルは、すべてフィードバックモジュール13にフィードバックされる。フィードバックモジュール13は、フィードバックされた情報ベクトルからプロファイル更新用データ、例えばユーザプロファイルを更新するためのテキストデータを作成するとともに、作成したプロファイル更新用データと予め作成されたプロファイル作成用データ（例えば出現頻度や情報種別等を表すテキストデータ）とを併合し、これに基づい

て該当するユーザプロファイルを更新する。このようにして、次のフィルタリングの際に、ユーザの関心を反映させるようにしている。

【0006】なお、上記フィルタリングモジュール11における個々の情報ベクトル a_i （ $i=1, 2, \dots, n$ ）とユーザプロファイル u との類似度は、下記(1)式に示されるように、両ベクトルのなす角 θ の余弦によって決定される。

【0007】

$$\text{【数1】} \cos \theta = a_i \cdot u / |a_i| \cdot |u| \quad (1)$$

【0008】但し、ベクトル x （ $x : a_i$ 、または u ）が (x_1, x_2, \dots, x_n) で表されるとき、ベクトルの絶対値 $|x|$ は $\sqrt{\sum x_i^2}$ で表される。 Σ は $k=1 \sim n$ までの総和である。(1)式に示されるように、 $\cos \theta$ は、情報ベクトル a_i とユーザプロファイル（ベクトル） u との内積の値を、各ベクトルの絶対値でそれぞれ除算することによって求められる。各情報ベクトル a_i とユーザプロファイル u とは、両者の角度が小さいほど類似性が強くなる。従って、類似性があるかどうかを判定するための基準となる角度を ψ とすると、フィルタリングモジュール11では、 $\cos \theta > \cos \psi$ 、即ち $\theta < \psi$ であれば、情報ベクトル a_i はユーザの関心に近い内容を有すると判定することができる。上記情報フィルタリング装置2は、この原理を利用したものである。

【0009】なお、ユーザプロファイルの生成方法については、TF/IDF法を代表とする方法が数多く提案されている（TREC: Text Retrieval Conference, SIGIR等）。また、プロファイルの更新については、Ide Dec-hi法、Ide regular法、Rocchio法等の多くの論文において、そのアルゴリズムが提案、引用されている（“WWW上の電子新聞に対する情報フィルタリングとその評価：菅井他、情処学会情報学基礎研究会、1996”、“The Effect of Adding Relevance Information in a Relevance Feedback Environment: C. Buckley他、SIGIR'94”や“Incremental Relevance Feedback for Information Filtering: J. Allan, SIGIR'96”等）。

【0010】

【発明が解決しようとする課題】上述のように、従来の情報フィルタリング装置では、ユーザが自分の関心と一致していると判定した情報ベクトルのすべてをフィードバックし、それに基づいてユーザプロファイルを更新している。しかし、ユーザは自分の関心を明確には意識していない場合が多く、ユーザが選別したベクトル集合 B が、実際にはそのユーザ自身の関心を的確に表していない場合がある。このため、情報ベクトルをフィードバックしてユーザプロファイルを更新しても、次のフィルタリングの際に必ずしもユーザの関心を正確に反映することができないという問題があった。また、従来の情報フィルタリング装置では、ユーザの関心の継続度合いを考慮していないために、ユーザが自分の関心外の情報を表

すベクトル集合を一時的に選別しまう場合があり、それが、そのままフィードバックされてユーザプロフィールが更新されてしまうため、フィルタリング精度が低下してしまうという問題があった。

【0011】そこで本発明の課題は、ユーザの関心がより正確に反映されるようにユーザプロフィールを更新するとともに、一時的に関心外の情報を表すベクトル集合が選別されても、それによる影響を最小限に抑えることが可能な情報フィルタリング技術を提供することにある。

【0012】

【課題を解決するための手段】上記課題を解決するため、本発明は、改良された情報フィルタリング方法を提供する。この方法は、コンピュータ装置を用いて情報フィルタリングを行う方法であって、複数の電子化情報をそれぞれ情報ベクトルとして表現し、これらの情報ベクトルのうち、予め定めた電子化情報に対するユーザの関心度合いを表す基準ベクトルとの類似度が第1基準値を越え、且つユーザが有用と判定した情報ベクトルの集合を第1ベクトル集合として特定する過程と、特定した第1ベクトル集合から互いの類似度が前記第1基準値より大きい第2基準値を越える情報ベクトルの集合を抜き出して第2ベクトル集合とし、この第2ベクトル集合に基づいて前記基準ベクトルを更新する過程とを含み、更新された基準ベクトルを次の第1ベクトル集合の特定に用いることを特徴とする。このような過程を経て情報フィルタリングを行うことで、更新される基準ベクトルは、ユーザの関心度合いを反映したものとなり、次回以降のフィルタリングの精度を高めることができるようになる。

【0013】本発明の方法では、第1ベクトル集合を特定した後、この第1ベクトル集合と過去に特定された第1ベクトルとの和集合から互いの類似度が前記第1基準値より大きい第2基準値を越える情報ベクトルの集合を抜き出して第2ベクトル集合とすることも可能である。この場合、過去に特定された第1ベクトル集合のうち、生成されてから所定期間を経過した情報ベクトルを除去したうえで第2ベクトル集合を抜き出すようにすることが好ましい。このようにすることで、ユーザが一時的に傾向の異なるベクトル集合を選別した場合でも、その影響を軽減させることができる。また、過去の選別結果による影響を抑えて、最新のフィルタリング時点でのベクトル集合の選別傾向を強く反映させることができるようになる。

【0014】本発明は、また、上記方法の実施に適した情報フィルタリング装置を提供する。本発明の情報フィルタリング装置は、それぞれ情報ベクトルとして表現された複数の電子化情報を入力するとともに複数の情報ベクトルのうち予め定めた電子化情報に対するユーザの関心度合いを表す基準ベクトルとの類似度が第1基準値を

越える情報ベクトルの集合を特定するフィルタリング手段と、このフィルタリング手段により特定された情報ベクトルの集合からユーザが有用と判定した第1ベクトル集合を選別する有用性判定手段と、前記第1ベクトル集合から互いの類似度が前記第1基準値よりも大きい第2基準値を越える情報ベクトルの集合を第2ベクトル集合として特定するベクトル分割手段と、前記第2ベクトル集合に基づいて前記基準ベクトルを更新するベクトル更新手段と、を備えて成る。

10 【0015】前記ベクトル分割手段は、例えば、第1ベクトル集合に属する個々の情報ベクトルについて互いの類似度が前記第2基準値を越える情報ベクトルの数を検出し、検出数が最大になるベクトル集合を前記第2ベクトル集合として特定するように構成される。このとき、前記検出数が同数となるベクトル集合が複数存在する場合は、時間的に新しい情報ベクトルをより多く含む一つのベクトル集合を第2ベクトル集合として決定するように構成することが望ましい。

【0016】本発明の情報フィルタリング装置において、前記第1ベクトル集合を一定時間蓄積するベクトル蓄積手段をさらに備えるようにしてもよい。この場合、前記ベクトル分割手段は、最新の第1ベクトル集合と前記ベクトル蓄積手段に蓄積されている第1ベクトル集合との和集合から前記第2ベクトル集合を特定するように構成する。これにより、ユーザの関心の継続度合いを考慮したフィルタリングが可能になる。

【0017】本発明は、さらに、上記情報フィルタリング方法ないし情報フィルタリング装置の機能をコンピュータ装置上で具現化するための情報フィルタリング用プログラムが記録された記録媒体をも提供する。このプログラムを読み取ったコンピュータ装置は、下記の処理を少なくともこの順に実行する。

30 (1) 複数の電子化情報をそれぞれ情報ベクトルに変換する処理、(2) 変換された複数の情報ベクトルのうち、予め定めた電子化情報に対するユーザの関心度合いを表す基準ベクトルとの類似度が第1基準値を越え、且つユーザが有用と判定した情報ベクトルの集合を第1ベクトル集合として特定する処理、(3) 特定した第1ベクトル集合から互いの類似度が前記第1基準値より大きい第2基準値を越える情報ベクトルの集合を抜き出して第2ベクトル集合とし、この第2ベクトル集合に基づいて次の前記第1ベクトル集合の特定に用いる前記基準ベクトルを更新する処理。

【0018】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態を詳細に説明する。図1は、本発明を適用した情報フィルタリング装置の一実施形態を表す機能ブロックと、1回目のフィルタリングの概要を示す図であり、図7に示した従来装置と同一機能の構成要素については同一符号を付してある。なお、図中、実線は処理の流

れ、破線はデータ類の流れを表す。

【0019】本実施形態の情報フィルタリング装置1は、コンピュータ装置が所定のプログラムを読み込んで実行することにより形成される、ユーザプロフィール格納部10、フィルタリングモジュール11、有用性判定モジュール12、ベクトル分割モジュール13、関心ベクトルデータベース(DB)14、フィードバックモジュール15を備えて構成される。上記プログラムは、通常、コンピュータ装置の内部記憶装置あるいは外部記憶装置に格納され、随時読み取られて実行されるようになっているが、コンピュータ装置とは分離可能な記録媒体、例えばCD-ROMやFD等に格納され、使用時に上記内部記憶装置または外部記憶装置にインストールされて随時実行に供されるものであってもよい。

【0020】この情報フィルタリング装置1における各部の動作を、電子化された新聞記事のフィルタリングを例に挙げて説明する。フィルタリング対象となる新聞記事は、情報ベクトルの集合、つまり従来装置の場合と同様、ベクトル集合Aに変換されており、図示しない入力手段を通じて、情報フィルタリング装置1に入力されるものとする。まず、図1のほか、図2(a)、(b)の原理説明図、図3のフィルタリング手順説明図をも参照して1回目のフィルタリングの内容を説明する。

【0021】図3に示すように、1回目のフィルタリングでは、ユーザが予め定めた関心を持つテーマに該当する記事の識別データ、出現頻度データ等を取得し、これを初期プロフィール作成用データとする(ステップS101)。そして、取得した初期プロフィール作成データから初期ユーザプロフィール u_0 を作成し、ユーザプロフィール格納部10に格納しておく(ステップS102)。次に、上記入力手段より入力された新規の新聞記事に関するベクトル集合Aを用意する。このベクトル集合Aに属する情報ベクトルは、それぞれ a_i ($i=1, 2, \dots, n$)で表される。フィルタリングモジュール11では、*

$$d(b_i, b_j) = b_i \cdot b_j / |b_i| \cdot |b_j| > \cos \gamma > \cos \psi \quad (3)$$

【0026】この情報ベクトル b_{\dots} は、ユーザが選別したベクトル集合Bのなかで、最も重要な情報を表す情報ベクトルとなる。そして、図2(a)に示されるように、情報ベクトル b_{\dots} に対して上記(3)式を満たすベクトルの集合(図示の網掛け部分)をフィードバック用ベクトル集合 $C'(T) = \{c'_1, c'_2, \dots, c'_n\}$ 、その他のベクトル集合(図示の網掛けのない部分)を待機用ベクトル集合 $C''(T) = \{c''_1, c''_2, \dots, c''_{n-1}\}$ とし、フィードバック用ベクトル集合 $C'(T)$ のみを抜き出す。なお、Tは情報ベクトルが作成されてからの時間情報であり、図2(a)におけるベクトル集合 Ω は、ユーザの関心だと判定したベクトル集合である。

【0027】ベクトル分割モジュール13より抜き出されたフィードバック用ベクトル集合 $C'(T)$ は、フィードバックモジュール15で初期プロフィール作成用デー

* 各情報ベクトルとユーザプロフィール格納部10からのユーザプロフィール u_0 との類似性を判定する(ステップS104)。具体的には、ユーザの関心との類似性を表すしきい値角(第1基準値)を ψ として、下記(2)式を条件を満たすかどうかを判定する。

【0022】

$$[\text{数2}] \quad a_i \cdot u_0 / |a_i| \cdot |u_0| > \cos \psi \quad (2)$$

【0023】条件を満たさない記事のベクトル集合については非類似、つまりユーザの関心外のものとして、該当する情報ベクトルを廃棄する(ステップS104: No, S105)。一方、条件を満たす記事のベクトル集合については類似する、つまりユーザが関心のある情報についてのベクトル集合と判定し、これらを有用性判定モジュール12に送る(ステップS104: Yes, ステップS106)。ユーザは、有用性判定モジュール12を通じて、上記ベクトル集合から実際に関心のある記事についてのベクトル集合Bを選別する(ステップS107)。ここまでの処理は、従来装置の場合と基本的には同じである。

【0024】本実施形態では、さらにベクトル集合Bをベクトル分割モジュール13に送り、ここでフィードバック用ベクトル $C'(T)$ と待機用ベクトル $C''(T)$ とに分割する(ステップS108)。ベクトル分割に際しては、まず、図2(a)、(b)に示されるように、ユーザが選択した個々の新聞記事の特徴を表す複数の情報ベクトル b_i のうち、互いの情報ベクトル b_i ($i \neq j$)のなす角の余弦と予め定めたしきい値角 γ ($\gamma \leq \psi$)とを比較し、下記(3)式の条件を満たす他の情報ベクトル b_j をより多く得ることができる情報ベクトル b_i を特定してこれを b_{\dots} とする。このときの $\cos \gamma$ が第2基準値となる。

【0025】

【数3】

と併合されてプロフィール作成データとなり(ステップS109)、これに基づいてユーザプロフィールの更新がなされる(ステップS110)。更新後のユーザプロフィールは u_1 となる。

【0028】分割されたフィードバック用ベクトル集合 $C'(T)$ と待機用ベクトル集合 $C''(T)$ は、関心ベクトル集合C(T)として、関心ベクトルデータベース14に格納され、それぞれ設定時間 τ だけ保持される。設定時間 τ を越えた情報ベクトルについては、随時廃棄されるようになっている。

【0029】次に、2回目(以降)のフィルタリングの内容を図4の機能ブロック、処理及びデータ類の流れを示した図4、原理図を示した図5、及びフィルタリング手順を示した図6を参照して説明する。

【0030】2回目(以降は)、既に得られた関心ベク

トル集合C(T)をも利用し、ユーザの関心の継続度を考慮したフィルタリングを行う。すなわち、図6に示すように、ユーザが予め定めた関心を持つテーマに該当する記事の識別データ、出現頻度データ等を取得し、これをプロファイル作成用データとする(ステップS201)。また、ユーザプロファイル格納部10からユーザプロファイルu_iを取得する(ステップS202)。そして、フィルタリングモジュール11において、新規の記事情報に関するベクトル集合Aに属する情報ベクトルと取得したユーザプロファイルu_iとの類似性を判定する(ステップS203)。類似するかどうかの条件は1回目のフィルタリングの場合と同様であり、条件を満たす記事のベクトル集合のみが有用性判定モジュール12に送られ、さらに、有用性判定モジュール12を通じて、ユーザが実際に関心のある新聞記事についてのベクトル集合Bのみが選別される(ステップS204~S207)。ここまでの処理は、1回目のフィルタリングの場合と同じである。

【0031】2回目(以降)のフィルタリングでは、関心ベクトルDB14に蓄積された関心ベクトル集合C(T)のうち、時間T≤設定時間τとなるベクトル全体を取得し、上記のようにして選別されたベクトル集合Bと併合させる(ステップS208)。そして、これにより得られた和集合から前述の情報ベクトルb_{...}を求めてベクトル再分割を行い、フィードバック用ベクトル集合を抜き出す(ステップS210)。なお、再分割の際に、情報ベクトルb_{...}の候補が複数存在した場合に、時間的に新しい情報をより多く含む情報ベクトルをb_{...}とする。また、このフィードバック用ベクトル集合と待機用ベクトル集合とを併合して関心ベクトルDB14に蓄積する。

【0032】フィードバックモジュール15では、プロファイル作成データと再分割により得られたフィードバック用ベクトル集合とを併合してプロファイル更新用データを作成する(ステップS210)。このとき、重複したデータについては、廃棄ないし重複を無視してデータを作成する。そして、作成されたプロファイル更新用データに基づいてユーザプロファイルu_iを更新する(ステップS211)。更新されたユーザプロファイルはu_iとなる。

【0033】このように、本実施形態のフィルタリング装置1は、ユーザプロファイルの更新に際して、情報ベクトルb_{...}の近傍に位置する情報ベクトルの集合、すなわちフィードバック用ベクトル集合C'(T)内の情報ベクトルのみをフィードバックするようにしたので、ユーザ自身が「自分の関心と合致する」と判定した情報ベ

クトルであっても、相互の関連性の強さが弱い情報ベクトルについては自動的にフィードバックの対象から外されるようになり、ユーザの関心が客観的且つより一層明確に反映されたユーザプロファイルが生成されるようになる。

【0034】また、2回目以降のフィルタリングでは、ユーザによって選択されたベクトル集合Bと、関心ベクトルDB14に蓄積された過去の関心ベクトル集合のうち設定時間τを越えない情報ベクトルとの和集合に基づいてフィードバック用ベクトルを得るようにしたので、過去数回の判定情報のうち、ユーザの関心が継続していると推定される情報に基づいてユーザプロファイルが更新されるようになり、何らかの要因によって一時的にユーザ関心外の情報が選別されたとしても、その影響を小さく抑えることができるようになる。

【0035】

【発明の効果】以上の説明から明らかなように、本発明によれば、ユーザにより選別された情報同士の間連性の強さとユーザの関心の継続性という2つの要素を考慮して情報ベクトルのフィードバックがなされるので、大量の電子化情報のなかからユーザの関心により合致した情報のみを自動的に獲得できるようになり、情報フィルタリング装置としての実用性が格段に高まる効果がある。

【図面の簡単な説明】

【図1】本発明の一実施形態に係るフィルタリング装置の機能ブロックと1回目のフィルタリングの概略を示す図。

【図2】(a)は分割時のベクトル集合の説明図、

(b)は個々の情報ベクトルの状態説明図。

【図3】1回目のフィルタリングの手順説明図。

【図4】本実施形態のフィルタリング装置における2回目のフィルタリングの概略を示す図。

【図5】本発明の原理を説明するための図で、分割時のベクトル集合の説明図。

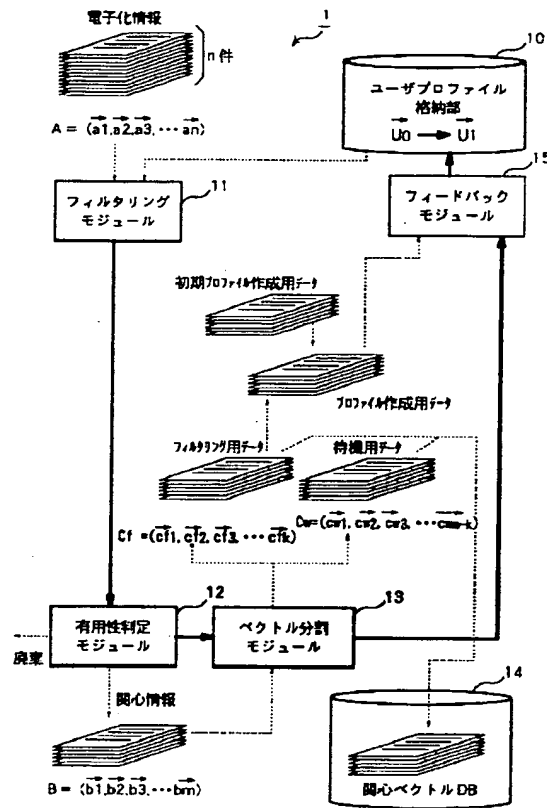
【図6】2回目以降のフィルタリングの手順説明図。

【図7】従来のフィルタリング装置の機能ブロックとフィルタリングの概略を示す図。

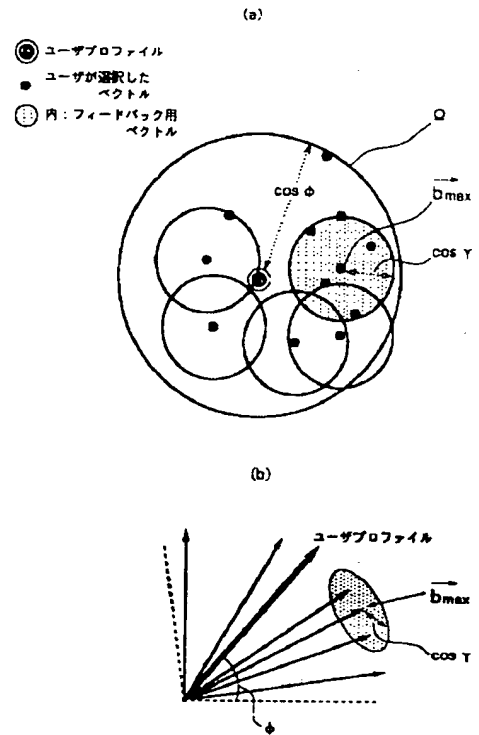
【符号の説明】

- 1、2 フィルタリング装置
- 10 ユーザプロファイル格納部
- 11 フィルタリングモジュール
- 12 有用性判定モジュール
- 13 ベクトル分割モジュール
- 14 関心ベクトルDB(データベース)
- 15 フィードバックモジュール

【図1】

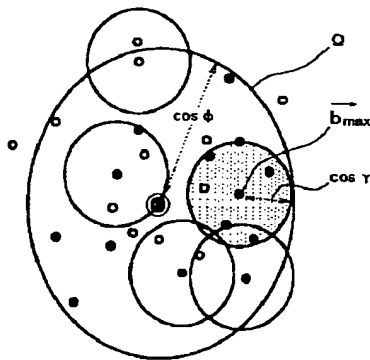


【図2】

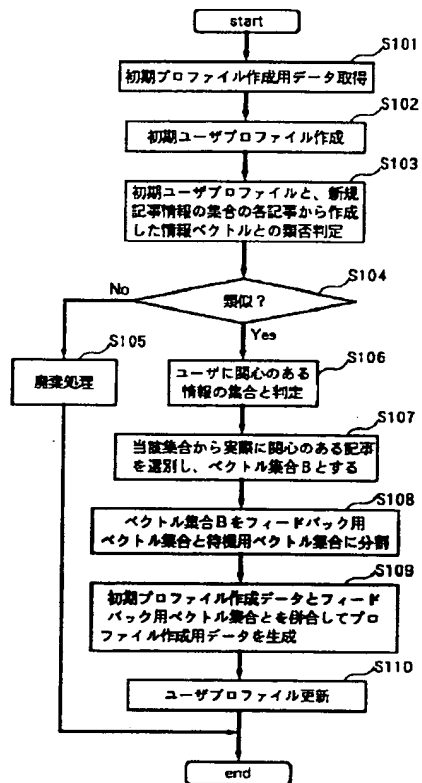


【図5】

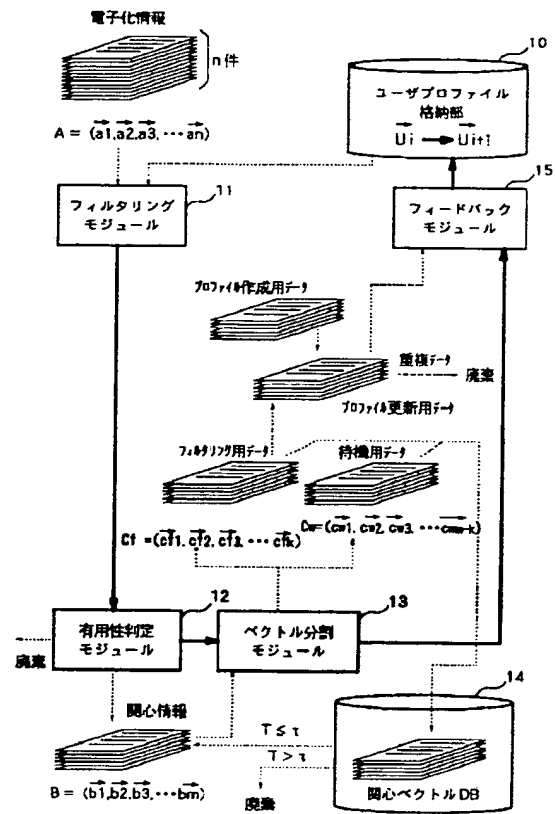
- ユーザプロフィール
- ユーザが選択したベクトル
- $T \leq 1$ を満たす待機用ベクトル
- 内: フィードバック用ベクトル



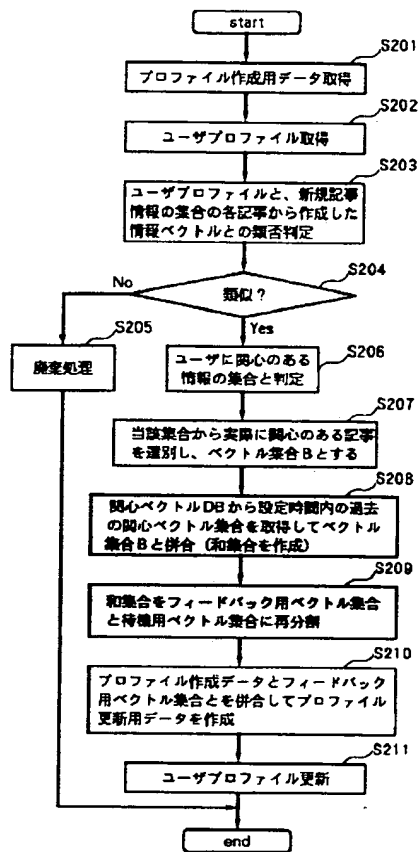
【図3】



【図4】



【図6】



【図7】

